DOCUMENT RESUME

ED 202 601 PS 012 203

AUTHOR Becker, Wesley C.; Gersten, Russell
TITLE A Follow Up of Follow Through: The Later Effects of the Direct Instruction Model on Children in Fifth and Sixth Grades.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Apr 81
GRANT OEG-007507234
NOTE 34p.; Paper presented at the Biennial Meeting of the Society for Research in Child Development (Boston, MA, April 2-5, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Arithmetic; Comparative Analysis; Decoding (Reading); Elementary Education; *Elementary School Students; Followup Studies; *Low Income Groups; Oral Language; Problem Solving; *Program Effectiveness; Small Group Instruction; Spelling
IDENTIFIERS *Direct Instruction Model; Metropolitan Achievement Tests; *Project Follow Through; Wide Range Achievement Test

ABSTRACT

The later effects of the Direct Instruction Follow Through program were assessed in five diverse sites in the United States. Low income fifth and sixth graders who had completed the full 3 years of this first through third-grade program were tested on the Metropolitan Achievement Test (MAT), Intermediate Level, and the Wide Range Achievement Test (WRAT). Scores were contrasted with those of children in local comparison groups using analysis of covariance procedures. Results indicated consistently strong, significant effects in WRAT Reading (decoding), consistent effects in MAT Math Problem Solving and Spelling, and moderate effects in most other academic domains. Students appeared to retain the knowledge and problem solving skills they had mastered in the primary grades. However, without a continuing program, most students demonstrated losses when compared to the standardization sample of the achievement tests. Implications for improved instruction in the intermediate grades are discussed. (Author/MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED202601

A Follow Up of Follow Through: The Later Effects of the Direct Instruction Model on Children in Fifth and Sixth Grades[1]

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Russell Gersten

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Wesley C. Becker

Russell Gersten

University of Oregon

Paper presented at Conference on a Society for Research on Child Development; Boston, Massachusetts, April, 1981

Running Head: Follow Up

PS012203

Abstract

The later effects of the Direct Instruction Follow Through program were assessed in five diverse sites. Low income fifth and sixth graders who had completed the full three years of this first-through-third-grade program were tested on the Metropolitan Achievement Test (Intermediate level) and the Wide Range Achievement Test. Results were contrasted with those of children in local comparison groups using analysis of covariance procedures. Results indicated consistently strong, significant effects in WRAT Reading (decoding), consistent effects in Math Problem Solving and Spelling, and moderate effects in most other academic domains. Students appeared to retain the knowledge and problem solving skills they have mastered in the primary grades. However, without a continuing program, most students demonstrated losses when compared to the standardization sample of the achievement tests. Implications for improved instruction in the intermediate grades were discussed.

Project Follow Through evaluated a variety of educational approaches to teaching low income children in various communities from kindergarten (or first grade) through third grade. The originators of Follow Through believed that gains made by students in Head Start could be enhanced and solidified in a comprehensive, systematic three or four year program. One of the approaches found to be most effective in the longitudinal evaluation conducted by Abt Associates and Stanford Research Institute under the auspices of USOE was the Direct Instruction Model (Bereiter & Kurland, in press; Kennedy, 1978; Stebbins, St. Pierre, Proper, Anderson & Cerva, 1977).

This model represents a highly structured approach to early childhood education with an emphasis on high levels of academic-engaged time through small-group instruction in reading, oral language, and arithmetic. The Distar® curriculum materials used in this approach are designed to explicitly teach general principles and problem-solving strategies. Teachers and paraprofessional aides are trained to teach these programs in a fast-paced, dynamic fashion with high frequencies of unison group responses and systematic corrections of student errors.

In some communities, Follow Through began in kindergarten and lasted four years; in others, Follow Through began in first grade and lasted three years. For a variety of logistical reasons (primarily having to do with difficulties in obtaining cooperation in large northern cities) this follow up study deals only with the three-year programs. (See Becker & Engelmann, 1978, for further details.)

Results of the national evaluation (Stebbins et al., 1977) indicated a high proportion of significant positive effects for both three- and four-year programs. When compared to a pooled national comparison group, Direct

Instruction students in three-year sites performed significantly higher than comparison students in 60 percent of the instances for Total Reading, 80 percent for Total Math, 100 percent in Language, and 50 percent in Spelling. Comparable percentages for the other sponsors with three-year programs are 23 percent in Reading, 21 percent in Math, 9 percent in Language, and 7 percent in Spelling.

The absolute level of performance was typically higher for students in the four-year programs than the three-year programs, particularly in Reading. Because of this fact, the results to be reported here represent a low estimate of what might have been achieved. Table 1 presents mean scores on all subtests of the Wide Range Achievement Test, Level 1 (WRAT) and 1970 Metropolitan Achievement Test, Elementary Level (MAT) (Durost, Bixler, Wrightstone, Prescott & Balow, 1970) for all low income students in the three-year program. These data were collected for four cohorts

---

Insert Table 1 about here

---

of children in eight communities. The WRAT scores, on the left, are presented in the form of a norm-referenced comparison (Horst, Tallmadge & Wood, 1975; Tallmadge, 1977). The mean standard scores at the beginning of the program (entry into first grade) and end of the program (end of third grade) are converted to percentiles in order to assess growth against the norm sample of the WRAT. The right half of Table 1 presents Metropolitan (MAT) scores. Again, the mean standard scores at the end of third grade were converted to a percentile. No pretest level of the MAT was available. The performance of Follow Through students is contrasted to typical third grade performance in reading and math for low income, minority students according to Office of Education reports (Ozenne et al.,

1974; 1976), and is represented by the jagged arrows. Note that MAT Math and Language scores are within a few percentile points of the national median, and all scores are significantly above the typical level of low income, minority students. The large discrepancy between decoding skills (word reading as assessed by the WRAT) and reading comprehension scores (assessed by MAT Reading) is likely due primarily to low income children's problems with the large, virtually uncontrolled reading vocabulary required on third grade achievement tests (which reflects the content of fourth grade textbooks; Becker, 1977). Schools expect vocabulary development to occur at home, and the evidence clearly indicates this is not happening in many low income homes. Note that no corresponding discrepancy appears in Math, where the MAT tests computation, high-order problem solving, and conceptual skills, and the WRAT tests only computation. It seemed important to examine the later effects of the program to see if the students maintained and built on the gains they made during the first three years of elementary school.

Design

A quasi-experimental design (Cook & Campbell, 1979) was utilized. Five Direct Instruction Follow Through projects agreed to participate. In each case, roughly equivalent comparison groups were located by the school district. Demographic information was collected on income level, race, sex, primary home language, and mother's education. These variables were used as potential covariates in the analyses. Despite the problems in using analysis of covariance (Campbell & Boruch, 1974; Elashoff, 1969; House, Glass, McLean & Walker, 1978), it was the only feasible option. As Cook and Campbell (1979) state, when one is using a quasi-experimental design and imperfect data analysis techniques, it is crucial to replicate. In this case, replications were conducted: (a) across five communities, (b) across two grade levels, and (c) across two cohorts of children.

Subjects and Testing

In the spring of 1975, fifth and sixth grade students who had previously completed the Follow Through program were tested on all subtests of the MAT (1970), Intermediate Level, and Levels I and II of the Reading subtest of the WRAT (Jastak & Jastak, 1965). The same battery was given to children in the local comparison group who had undergone traditional education in the community. The only selective factors operating in choosing sites were the availability of a local comparison group and district cooperation. The five sites which agreed to participate were Dayton, Ohio; East St. Louis, Illinois (representing urban Black populations); Tupelo, Mississippi (with a rural Black population); Smithville, Tennessee (with a rural Anglo population); and Uvalde, Texas (with a primarily Hispanic population).[2] In 1976, the study was replicated in four of the five sites. Dayton did not participate in the replication study.

The 1975 study involved 624 Follow Through graduates and 567 non-Follow Through students; the 1976 study, 473 Follow Through graduates and 403 non-Follow Through children. Table 2 presents sample sizes and demographic information for each site. Low income students were sought in each case. As it turned out, there were a few non-low income students in most samples, so income level was used as a potential covariate.

Insert Table 2 about here

Testing was conducted by local staff after training by University of Oregon supervisors on the procedures specified in the publishers' test manuals. Actual testing was monitored by supervisors to ensure that standard proce-

dures were followed. In Tupelo, Mississippi, the California Achievement Test results (collected by the local district) were used in place of the MAT and the reading scores were converted to MAT equivalents using the Anchor Test Study (Loret, Seder, Bianchini & Vale, 1974).

Strategies for Data Analysis

Analysis of covariance using ethnicity, income, home language, mother's education, sex, and number of siblings was performed on each subtest at each site. One strategy that was considered, and ultimately rejected, was pooling together all Follow Through and all comparison (non-Follow Through) children in all five sites (cf. Goodrich & St. Pierre, 1979). Despite the immense gain in sample size and statistical power, this option was deemed inappropriate since subjects in both Follow Through and comparison groups came from at least four highly distinct populations (urban Black, rural Black, rural Anglo, and Hispanic) and it seemed highly unlikely that the assumptions of ANCOVA would be met. Thus, separate ANCOVAs were performed on each site for each subtest for each year.

The remaining problem was one of meaningfully synthesizing and collating the results of the multitudinous analyses. Three strategies were adopted. Using the site as a unit of analysis was rejected because there were too few sites to give adequate power to any test.

In the first analysis, the results of each ANCOVA were classified as: (a) significant ($p < .05$, two-tail), (b) suggestive of a trend ($.15 > p > .05$), or (c) non-significant. For each subtest, the number of sites falling into each category was tabulated. Since sample sizes tended to be small (ranging from 25 to 117, but averaging around 50), it seemed appropriate to record those sites in which $.15 > p > .05$. The researchers reasoned that, if, for a particular subtest (e.g., Math Problem Solving), only 1 out of 10 findings

was positive at .15 > p >.05, this would rightfully be considered a chance finding. If, on the other hand, 8 of the 10 comparisons were found to be "suggestive" at the .15 level, this would be evidence of a replicable phenomena.

The second procedure used was the meta-analysis technique advocated by Gage (1977) (after Jones and Fiske, 1953). For each site level ANCOVA the exact p values (both significant and nonsignificant) are converted to chi-square ratios (with two degrees of freedom). Total chi-square values were then tested for significance with 2(n-1) degrees of freedom (where n = number of studies in the meta-analysis). This technique is the one of the only meta-analytic techniques to offer statistical significance levels for the comparisons.

Finally, the average magnitude of effect in pooled standard deviation units for each subtest at each grade level was calculated (Glass, 1976; Pillmer & Light, 1980; Smith & Glass, 1977). The method gives an estimate of the treatment effect that is not biased by the differential sample sizes at the various sites.

Results

Table 3, which summarizes the data from the first method of analysis, indicates the number of site level ANCOVAs comparisons falling into each of the following categories ($p < .05$; $.05 < p < .15$; ns) for both the 1975 and 1976 studies. Table 4 presents the results of the Gage (1977) and Jones and Fiske (1953) meta-analysis procedures; chi-square values are shwon for grade 5, for grade 6, and then for the total number of comparisons. Table 5, which presents the mean magnitude of effect in pooled SD units (Glass, 1976) contains results of the third method of analysis.

Insert Tables 3, 4, 5 about here

The strongest, most consistent, finding is for reading decoding, as assessed by both Level I and Level II of WRAT Reading: $\chi^2_{(30)} = 134.1$, $p < .005$ for Level I; $\chi^2_{(32)} = 134.4$, $p < .005$ for the Level II test. The mean magnitude of effects (in Table 5) range from .38 to .56 pooled standard deviation units, well over the conventional criteria set for educational significance (Horst et al., 1975; Stebbins et al., 1977). This test measures children's ability to accurately read words in isolation. The consistency is demonstrated across sites, grade levels, and levels of the test. The teaching of decoding (or word attack skills) is one of the strongest early outcomes of the Direct Instruction Model; mean end-of-third grade performance corresponds to the 67th percentile on the WRAT for entering-first grade students. It appears that these skills are maintained two and three years after the program ends.

A strong, consistent effect is also found on MAT Spelling (Table 4), with significant effects for both grades 5 ($p < .05$) and 6 ($p < .05$) and the combined sample ($p < .005$). It is possible that the enduring effect in spelling is related to the phonic and word attack skills the students mastered in the early grades.

The other strong, consistent effect is in Math Problem Solving. Table 3 shows that in each year, half of the site level ANCOVAs significantly favor Follow Through at either the .05 or .15 level, and no comparisons significantly favor the comparison children. The chi-square analyses (Table 4) are also significant at the .05 level for both Grade 5, Grade 6, and at the .005 level for the combined sample. Mean magnitude of effects are .27 for Grade 5 and .18 for Grade 6. Note that the Math Problem Solving effects are consistently stronger than Math Computation. At first, this would seem unusual for a program with a heavy emphasis on acquisition of basic skills. Yet, the finding is consistent with the emphasis in the Distar Arithmetic programs on teaching general-case, problem-solving strategies, including basic algebraic principles.

More variable effects are found for the Word Knowledge, Math Concepts, Science, Language subtests as well as the composite scores for Total Reading and Total Math. For example, for Word Knowledge, and Science the chi-square analyses (Table 4) indicate significant effects for the combined sample ($p < .01$) and the fifth grade ($p < .05$), but not sixth grade. Yet the magnitude of effects (Table 5) is actually somewhat higher for Grade 6. Site level analyses do not shed any great light on these patterns, other than indicating that the largest site, Uvalde, seemed to have consistent effects in Science and Word Knowledge (See Becker & Engelmann, 1978, and Becker & Gersten, 1979 for detailed site level analyses.)

Overall, there is reasonable evidence of significant later effects. Of the total of 180 comparisons in Table 3, 56 favor the Follow Through sample at the .05 level; none favor the comparison group. If one uses the more liberal .15 significance level to explore potential trends (Carver, 1978), 76 effects (42 percent) favor Follow Through, and only two (less than 1 percent) favor the comparison groups. The tests with the strongest effects are WRAT Reading, MAT Spelling, and MAT Math Problem Solving. Meta-analysis techniques revealed a similar pattern. The mean magnitude of effect (Table 5) is well above .33 pooled SD units for all levels of WRAT Reading and in the .17 to .26 range for most MAT subtests. The two tests with consistently low magnitudes of effect are Math Computation and Reading (comprehension). Using the Jones-Fiske (1953) statistical tests for the combined sample, there is evidence of significant enduring effects in all domains except MAT Reading (a test of reading comprehension). There are no consistent patterns indicating that one particular site or one particular grade level displayed more lasting effects. Thus, the effects appear to be due to the model.

To round out the picture, Table 6 presents the unadjusted percentiles for WRAT Reading and Total Reading, Total Math, Spelling, and Science on the MAT for each site in the 1975 study. These percentiles were converted

Insert Tables 6 and 7 about here

from the unadjusted mean standard scores. Table 7 presents the mean magnitude of effects for adjusted scores on a site level basis for the children who were fifth grade in 1975 and sixth grade in 1976. Table 7 shows reasonably high consistency across grade levels when the same children are followed.

Longitudinal Tracebacks: Gains and Losses of Follow Through Children Against the National Norm Sample

Table 8 traces the growth of the Follow Through children against the norm sample of the WRAT for the children during the three years of the Follow Through program, followed by their decline during the intermediate grades. East St. Louis is omitted since data were unavailable. Entry

Insert Tables 8 about here

(pretest) scores were unavailable from Tupelo and Dayton; however, on the basis of similar Follow Through children tested at entry in these communities in later years, one can estimate the entry scores at approximately the 14th percentile for Dayton and the 9th percentile for Tupelo (Becker & Engelmann, 1978).

Note in Table 8 that the major growth in reading decoding occurs during the first two years of school, when a major program emphasis is on word attack

skills. The level is basically maintained in grade 3. By grades 5 and 6, there are appreciable drops in Smithville and Tupelo. Although the Follow Through students significantly outperform non-Follow Through students at all sites on the WRAT, they are "losing" when compared to the norm sample.

This decline is even more dramatic on the Metropolitan Achievement Test (Table 9). The MAT is a well-normed comprehensive test of reading, math, and language, including both basic skills and higher-order cognitive operations (Bereiter & Kurland, in press; Wolf, 1978). To save space, Table 9 presents only percentile equivalents for the composite scores--Total Reading, Total Math, and Spelling. (The mean scale scores, standard deviations, and sample sizes, and detailed site level analyses are available in Becker and Engelmann, 1978, and Becker and Gersten, 1979.)

Insert Table 9 about here

By the end of the third grade, and three years of Direct Instruction Follow Through, all sites were within a few percentile points of the national median in Total Math and within one-fourth standard deviation in Spelling. In Total Reading, Smithville students are at the national median, the Dayton sample within one-fourth of a standard deviation unit, and Tupelo and Uvalde one-half standard deviation below. Of primary importance is that two years after the program ends, all samples make appreciable, significant drops against the national norm samples in both Math and Reading. In some instances there are even further drops during Grade 6. Although in many domains, Follow Through graduates outperform the comparison students in Grades 5 and 6, both Follow Through and non-Follow Through low income students are losing against the national norm sample. The losses are much smaller in MAT Spelling and WRAT Reading (decoding) than MAT Reading and Language.

Conclusions

From a methodological point of view, the three techniques used to summarize the results all yield a roughly comparable picture, though minor discrepancies emerged. These fluctuations are due to the fact that the Jones-Fiske (1953) chi-square technique weighs the larger site, Uvalde, with sample sizes of approximately 100, more heavily than the others, whereas the Glass (1976) magnitude of effects method weighs each site equally. The rationale for use of the former procedure are the basic tenets of statistical inference (e.g., Hays, 1973), that, all other conditions being equal, the larger the sample size, the more reliable the estimate of the treatment effect.

There are two basic findings to this study. The first is that there is evidence that in most domains assessed by standardized achievement tests, low income graduates of a three-year Direct Instruction Follow Through program perform better than comparable children in comparison groups in their communities who did not attend the program. These enduring effects are strongest and most consistent in WRAT Reading and Decoding, Math Problem Solving and Spelling. There are lesser effects in MAT Science, Math Concepts, Math Computation, and Word Knowledge. The fact that this study was conducted at five quite diverse sites across two grade levels and replicated in 1976 adds confidence to the results. Since none of the outcomes significantly favored the comparison groups at the .05 level, and 31% favored Follow Through, it is extremely unlikely that these results are due to chance. These results are from the second and third cohorts of Follow Through children, when the program was not fully developed. Also, only sites with a three-year (rather than four-year) program participated. For both these reasons, the results cited here, in all likelihood, represent a low

estimate of program effectiveness. For example, a quasi-experimental follow up study by Gutkin, Gersten, and Meyer (Note 3) demonstrated that low income fifth graders from the later cohorts of the four-year program in New York City were significantly outperforming comparison group children, and were well above national median levels in reading on the Comprehensive Test of Basic Skills (CTBS). Also, Weber and Fuhrmann (Note 4) reported significant later effects in reading and math on the California Achievement Test for ninth graders who had completed the program; Follow Through graduates were .24 standard deviation units (or .8 grade equivalent units) ahead of comparable students in the district.

The second finding is less optimistic. When compared to the national norm sample, these children invariably lose ground in the three years after they leave Follow Through.

Two reasonable conclusions can be formed from these findings. The first is that if students learn skills and problem solving strategies well, they do not lose this knowledge. Follow Through graduates often perform significantly higher than other low income fifth and sixth graders in their communities. The second conclusion is that without effective instruction which continues to build on these skills in the intermediate grades, the children are likely to lose ground against their middle-income peers. They are failing to master new computational skills (such as long division and complex multiplication); they are failing to develop higher-order comprehension skills, and failing to develop their vocabularies at the rate of non-low income students. Limited English-speaking students appear to lose the most. In order for these children to become fully literate adults, there appears to be a need for high quality instructional programs in the intermediate grades (and probably beyond). Key areas for program development are instruction in reading comprehension

(Jenkins, Stein & Osborn, in press; Perfetti & Lesgold, 1979; Resnick , in press); vocabulary development (Becker, 1977): independent study skills (Adams, 1980; Chall, 1979; Durkin, 1978-79); mathematical problem solving (Silbert, Carnine & Stein, 1981): expressive writing (Frederiksen, Whiteman & Dominic, 1981); and independent reading for information and pleasure (Brown & Smiley, 1977). An ever-increasing body of knowledge has accrued from descriptive and correlational studies of effective classroom practices in the elementary grades (Brophy & Evertson, 1976; Fischer et al., 1980; Gersten, Carnine & Williams, in press; Good & Beckerman, 1978; Stevens & Rosenshine, in press). These studies isolate teaching practices which are consistently effective, such as high student success rate, clarity of tasks, amount of guided practice and systematic use of correction procedures. There is now a need to implement and evaluate instructional programs in the intermediate grades that systematically utilizes principles of direct instruction, which include mastery learning, high levels of feedback, and incremental steps to develop independent reading, writing, and critical thinking.

Reference Notes

1. Gutkin, J., Gersten, R., & Meyer, L. Evaluation of P.S. 137 (New York City) Follow Through. Final report submitted to Joint Dissemination Review Panel, Department of Education, Washington, D.C., 1981.

2. Weber, B., & Fuhrmann, M. A study of District A's former Follow Through students' retention of basic skills after six years out of the program. Paper prepared for District A, April, 1978.

References

Adams, A. Use of direct instruction to teach an independent study method to skill deficient fifth grade students. Unpublished doctoral dissertation, University of Oregon, 1980.

Becker, W. C. Teaching reading and language to the disadvantaged--What we have learned from field research. Harvard Educational Review, 1977, 47, 518-543.

Becker, W. C., & Engelmann, S. Analysis of achievement data on six cohorts of low income children from 20 school districts in the University of Oregon Direct Instruction Follow Through Model. Follow Through Project, Technical Report #78-1. Eugene, OR: University of Oregon, 1978.

Becker, W. C., & Gersten, R. M. Follow up study of fifth and sixth graders: The 1976 replication study. Follow Through Project, Technical Report #78-1. Eugene, OR: University of Oregon, 1979.

Bereiter, C., & Kurland, M. A constructive look at Follow Through results. Interchange, in press.

Brophy, J., & Evertson, C. Learning from teaching: A developmental perspective. Boston: Allyn and Bacon, 1976.

Brown, A. L., & Smiley, S. S. Rating the importance of structural units of prose passages: A problem of metacognitive development. Child Development, 1977, 48, 1-8.

Campbell, D. T., & Boruch, R. F. Making the case of randomized assignments to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. A. Bennett & A. A. Lumsdaine (Eds.), Evaluation and experiment: Some critical issues in assessing social programs. New York: Academic Press, 1975.

Carver, R. P. The case against statistical significance testing. Harvard Educational Review, 1978, 48, 378-399.

Chall, J. S. The great debate: Ten years later, with a modest proposal for reading stages. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol 1). Hillsdale, NJ: Lawrence Erlbaum Associates, 1979.

Cook, T., & Campbell, D. T. Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally, 1979.

Durkin, D. What classroom observations reveal about reading comprehension instruction. Reading Research Quarterly, 1978-79, 15, 481-533.

Durost, W. N., Bixler, H. H., Wrightstone, J. W., Prescott, G. A., & Balow, I. H. Metropolitan Achievement Tests. New York: Harcourt Brace Jovanovich, 1970.

Elashoff, J. D. Analysis of covariance: A delicate instrument. American Educational Research Journal, 1969, 6, 383-401.

Fischer, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. Teaching behaviors, academic learning time and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), Time to learn. Washington, D.C.: USOE/NIE Printing, 1980.

Frederiksen, C. H., Whiteman, M., & Dominic, J. (Eds.) Writing: The nature, development, and teaching of written communication. Hillsdale, NJ: Lawrence Erlbaum, Associates, 1981.

Gage, N. The scientific basis of the art of teaching. New York: Columbia Teachers College Press, 1977.

Gersten, R., Carnine, D., & Williams, P. Measuring implementation of a structured educational model in an urban setting: An observational approach. Educational Evaluation and Policy Analysis, in press.

Glass, G. V. Primary, secondary, and meta-analysis of research. Educational Researcher, 1976, 5, 3-8.

Good, T. L., & Beckerman, T. M. Time on-task: A naturalistic study in sixth grade classrooms. Elementary School Journal, 1978, 78, 193-201.

Goodrich, R., & St. Pierre, R. Opportunities for studying later effects of Follow Through. Cambridge, MA: Abt Associates, 1979.

Hays, W. L. Statistics for the social services (2nd ed.). New York: Holt, Rinehart & Winston, 1973.

Horst, D. P., Tallmadge, G. K., & Wood, C. T. A practical guide to measuring project impact on student achievement. Monograph No. 1 on Evaluation in Education. Washington, D.C.: U.S. Government Printing Office, 1975.

House, E. R., Glass, G. V., McLean, L. D., & Walker, D. E. No simple answer: Critique of the "Follow Through" evaluation. Harvard Educational Review, 1978, 48, 128-160.

Jastak, J., & Jastak, S. Wide Range Achievement Test. Wilmington, Delaware: Jastak Associates, 1965.

Jenkins, J. R., Stein, M., & Osborn, J. What next after decoding: A look into instruction and research in reading comprehension. Exceptional Education Quarterly, in press.

Jones, L., & Fiske, D. W. Models for testing the significance of combined results. Psychological Bulletin, 1953, 50, 375-381.

Kennedy, M. Findings from the Follow Through planned variation study. Educational Researcher, 1978, 7, 3-11.

Loret, P. G., Seder, A., Bianchini, J. C., & Vale, C. A. Anchor test study: Equivalence and norms tables for selected reading achievement tests (grades 4, 5, 6). Office of Education Report No. 74-305. Washington, D.C.: U.S. Government Printing Office, 1974.

Ozenne, D., et al. United States Office of Education. Annual evaluation report on programs administered by the U.S. Office of Education, FY 1973. Washington, D.C.: Capital Publications. Educational Resources Division, 1974.

Ozenne, D., et al. United States Office of Education. Annual evaluation report on programs administered by the U.S. Office of Education, FY 1975. Washington, D.C.: Capital Publications. Educational Resources Division, 1976.

Perfetti, C. A., & Lesgold, A. M. Coding and comprehension in skilled reading and implications for reading instruction. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Pillemer, D. G., & Light, R. J. Synthesizing outcomes: How to use research evidence from many studies. Harvard Educational Review, 1980, 50, 176-197.

Resnick, L. B. Instructional psychology. Annual Review of Psychology, in press.

Silbert, J., Carnine, D., & Stein, M. Direct instruction arithmetic. Columbus, OH: Charles E. Merrill, 1981.

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. American Psychologist, 1977, 32, 752-760.

Stebbins, L., St. Pierre, R. G., Proper, E. L., Anderson, R. B., & Cerva, T. R. Education as experimentation: A planned variation model (Vols. IVA-D). Cambridge, Mass.: Abt Associates, 1977.

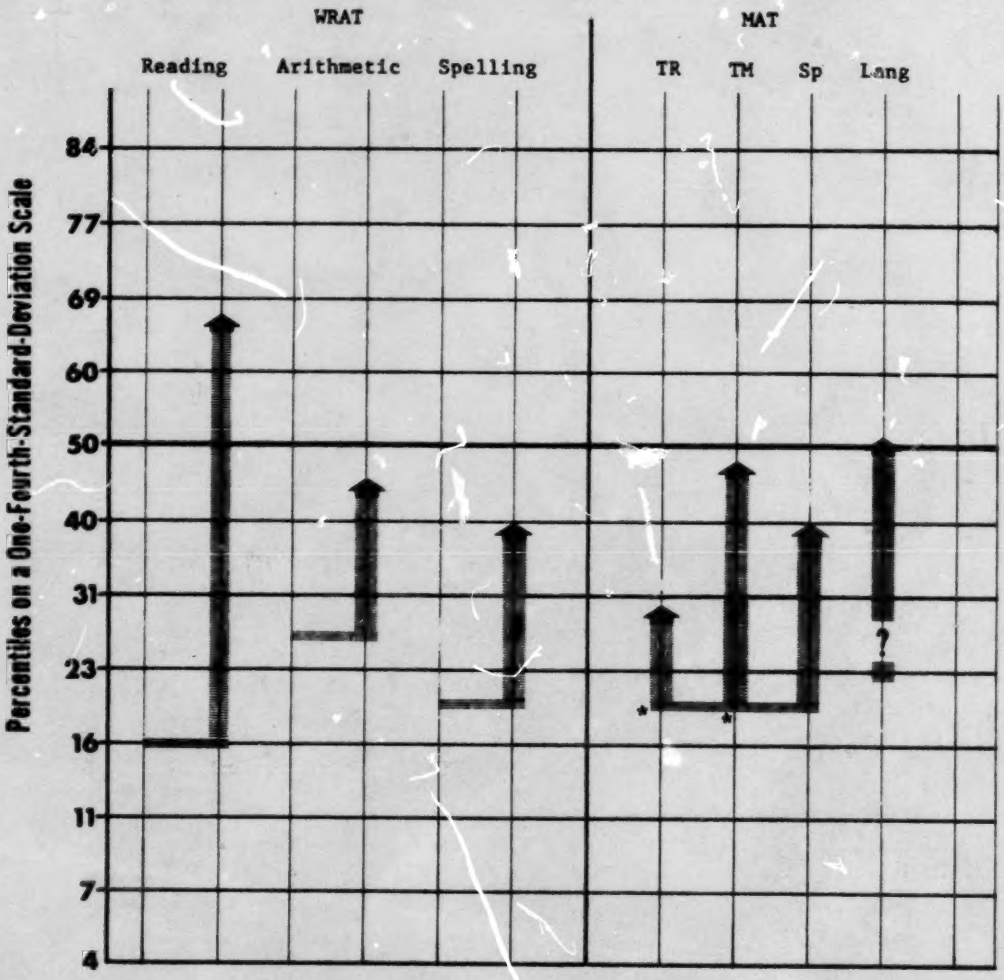
Stevens, R., & Rosenshine, B. Advances in research on teaching. Exceptional Education Quarterly, in press.

Tallmadge, G. K. JDRP ideabook. Washington, D.C.: USOE/NIE, 1977.

Wolf, R. Review of Metropolitan Achievement Test. In O.K. Buros (Ed.), The sixth mental measurements yearbook. Highland Park, NJ: The Gryphon Press, 1978.

Table 1

Norm-Referenced Gains on the WRAT and End of Third Grade Scores on MAT
Low-Income, Direct Instruction Follow Through students (Three-Year Program)



PRE	1		1		1					
POST		3		3		3	3	3	3	3
N	2418	4100	2451	4112	2428	4068	2392	2385	2410	2402
%-tile	16	67	26	45	20	40	30	48	40	51
S. S.	84.9	106.3	90.5	98.0	87.4	96.4	55	70.0	61.0	70.7
S. D.	14.4	19.0	13.7	9.7	15.1	15.4	9.0	11.0	11.0	12.3
G. E.	.51	4.6	.77	3.8	.64	3.6	2.9	3.7	3.3	4.1

*Typical end of third grade performance of low income minority students (Ozenne et al., 1976).

Table 2

Comparison of FT and NFT Groups on Selected Demographic Variables for the Fifth Grade

Site		Sample Size		Mother's Ed. Scale[a]		Percentage Low Income		Percentage Non-White	
		FT	NFT	FT	NFT	FT	NFT	FT	NFT
E. St. Louis IL	1975	43	45	4.35	3.68	.93	.98	1.00	1.00
	1976	45	44	4.76	-	1.00	-	1.00	-
Smithville TN	1975	47	51	3.30	3.70	.87	.54	.17	.00
	1976	71	38	3.43	3.40	.86	.86	00	.05
Uvalde TX	1975	117	86	2.86	2.34	.87	.95	1.00	.95
	1976	103	74	2.77	2.35	.98	.98	.98	.80
Dayton[a] OH	1975	104	87	4.75	5.00	.85	.71	1.00	.91
	1976[b]								
Tupelo MS	1975	46	35	3.22	3.50	.95	.97	.93	.53
	1976	56	42	4.11	4.15	.87	.95	.95	.64

[a] 5 = High school graduates

[b] Did not participate in 1976 study

- Missing data

Table 3. Significance Levels (two-tailed test) for Fifth and Sixth Grade Follow Through/Non-Follow Through Comparisons with Covariance Adjustment

	WRAT Reading		Metropolitan Achievement Test										Summary of Effects
			Reading			Math							
	I	II	Word Know	Rdg	Tot	Comp	Conc	Prob Solv	Tot	Sci	Lang	Spell	Total
1975 Study													
oring FT ($p < .05$)	7	7	3	2	2	2	3	3	3	2	2	1	37
FT ($.15 > p > .05$)	0	0	0	1	1	2	0	1	0	0	0	4	9
ns ($p > .15$)	1	1	7	7	7	4	5	4	5	6	6	3	56
oring NFT	0	0	0	0		0	0	0	0	0	0	0	0
al	8	8	10	10	10	8	8	8	8	8	8	8	102
1976 Study													
oring FT ($p < .05$)	5	4	0	1	1	0	2	0	0	1	3	2	19
FT ($.15 > p > .05$)	0	2	1	0	1	1	0	3	0	1	0	2	11
ns ($p > .15$)	2	2	7	7	6	1	4	3	5	4	3	2	46
oring NFT ($.15 > p > .05$)	0	0	0	0	0	1	0	0	1	0	0	0	2
al	7	8	8	8	8	3	6	6	6	6	6	6	78

Table 4. Pooled Results from 1975 and 1976 Follow Up Studies in Seven Follow Through Sites, Chi-Square Ratios Calculated by Jones & Fiske (1953) Meta-Analysis Procedures

Metropolitan Achievement Test (Intermediate) (1970)

Subtest	Grade 5			Grade 6			Combined		
	Chi-square	df	p	Chi-square	df	p	Chi-square	df	p
Word Knowledge	33.8	18	<.05	26.7	18	ns	60.5	36	<.01
Reading	30.3	18	<.05	20.3	18	ns	50.6	36	ns
Total Reading	36.2	18	<.05	22.7	18	ns	58.9	36	<.01
Language	22.7	14	ns	24.8	15	<.05	47.5	28	<.05
Spelling	29.1	14	<.05	27.2	14	<.05	56.3	28	<.005
Math Comp.	16.3	14	ns	25.5	14	<.05	41.8	28	<.05
Math Concepts	21.1	14	ns	33.3	14	<.05	54.3	28	<.01
Math Problem-Solving	33.2	14	<.05	28.0	14	<.05	61.2	28	<.005
Total Math	16.8	14	ns	24.8	14	<.05	41.6	28	<.05
Science	26.6	14	<.05	22.4	14	ns	49.0	28	<.01

WRAT Reading (Decoding)

Level I	80.0	14	<.005	54.0	16	<.005	134.1	30	<.005
Level II	73.3	16	<.005	61.1	16	<.005	134.4	32	<.005

Table 5. Mean Magnitude of Effects in Pooled Standard Deviation Units Between FT and NFT Samples- Pooled from 1976 and 1976 Studies

Test	5th Grade	6th Grade
WRAT Level II	.50	.51
WRAT Level I	.56	.38
MAT Word Knowledge	.19	.23
MAT Reading	.16	.14
MAT Total Reading	.20	.19
MAT Language	.21	.20
MAT Spelling	.24	.17
MAT Math Computation	.09	.13
MAT Math Concepts	.18	.24
MAT Math Problem Solving	.27	.18
MAT Total Math	.18	.26
MAT Science	.20	.26

Table 6. Unadjusted Percentiles for Follow Through and Non-Follow Through in 1975 Study

			MAT							
	WRAT Reading[a]		Total Reading		Total Math		Spelling		Science	
Grade 5	FT	NFT	FT	NFT	FT	NFT	FT	NFT	FT	NFT
E. St. Louis	--	--	17	17	16	18	41	39	11	14
Smithville	63	39	34	34	49	56	39	39	36	36
Uvalde	45	23	16	17	24	21	37	37	22	19
Dayton	61	27	20	16	19	12	34	27	20	13
Tupelo	42	27	18	18	--	--	--	--	--	--
Grade 6										
E. St. Louis	--	--	32	18	33	21	53	43	20	12
Smithville	73	25	36	21	52	28	59	39	42	26
Uvalde	32	18	15	15	22	17	39	28	21	17
Dayton	50	27	22	22	19	17	13	25	23	21
Tupelo	44	13	36	28	--	--	--	--	--	--

[a] Level I for Grade 5, Level II for Grade 6

Table 7. Mean Magnitude of Effects

in Pooled Standard Deviation Units

A Within-Cohort Follow Up of 5th Graders

in 1975 who were 6th Graders in 1976

Site	Uvalde		Smithville		E. St. Louis		Tupelo	
Date	1975 → 1976		1975 → 1976		1975 → 1976		1975 → 1976	
N	117	108	46	38	43	37	56	55
TEST: WRAT Reading Level II	.47	.41	.61	.53	x	-.15	.78	.46
WRAT Reading Level I	.47	.52	.52	.48	x	-.15	.63	-.37
MAT Word Knowledge	.03	.19	.06	.02	.00	-.03	.38	.18
MAT Reading	.21	.20	.25	.02	-.12	-.20	.20	.01
MAT Total Reading	.11	.20	.15	.02	.01	-.11	.28	.09
MAT Language	.13	.39	.09	.03	.03	-.06	x	x
MAT Spelling	.01	.25	.36	.37	.13	-.06	x	x
MAT Math Computation	.11	.04	.09	.15	.05	-.30	x	x
MAT Math Concepts	.10	.31	.09	-.08	.21	-.30	x	x
MAT Math Problem-Solving	.33	.23	.20	-.28	.02	-.22	x	x
MAT Total Math	.10	.16	.15	-.08	.01	x	x	x
MAT Science	.15	.36	.23	-.04	-.15	x	x	x

x Missing data

Table 8. Longitudinal Analysis of WRAT Reading at Entry and at the End of Grades 1, 2, 3, 5, and 6

	WRAT Reading (Level I)					
Grade	Pre-1 (Fall,1970)	1(1971)	2(1972)	3(1973)	5[a](1975)	6[b](1976)
Site						
Uvalde, TX						
Percentile	9	43	64	55	45	42
Mean Standard Score [c]	79.7	97.4	105.5	102.0	98	97
N	81	110	110	110	117	108
Smithville, TN						
Percentile	20	66	74	75	63	39
Mean SS	87.6	106.0	109.4	110.1	105	96
N	40	46	43	46	46	38
Tupelo, MS						
Percentile	--	39	57	55	42	34
Mean SS	--	96.0	102.5	102.1	97	94
N	--	24	25	25	56	55
Dayton, OH						
Percentile	--	52	73	67	61 [a]	--
Mean SS	--	100.7	108.8	106.7	104	--
N	--	97	104	102	93	--

[a] Estimate based on Raw Score for Level I

[b] Estimate based on raw score for Level II

[c] Mean = 100, SD = 15

- Missing data

Table 9

Longitudinal Analysis of Metropolitan Achievement Test

Same Children Followed from Third Through Sixth Grades

Total Reading Percentiles [a,b]

Site	Grade 3 (1973)	Grade 5 (1975)	Grade 6 (1976)
Uvalde, TX	31st	16th	16th
Smithville, TN	52nd	34th	26th
Tupelo, MS	28th	18th[c]	17th[c]
Dayton, OH	40th	20th	---

Total Math Percentiles [a]

Site			
Uvalde, TX	53rd	24th	19th
Smithville, TN	78th	49th	36th
Dayton, OH	55th	19th	---

Spelling Percentiles [a]

Site			
Uvalde, TX	40th	37th	35th
Smithville, TN	62nd	39th	39th
Dayton, OH	40th	34th	---

[a] Percentiles converted from mean standard score for the sample

[b] Third grade tested on Elementary Form, Grades 5 and 6 on Intermediate Form

[c] CAT scores converted by Anchor Study tables

--- Unavailable

Footnotes

1. This research was supported in part by U.S. Office of Education Grant No. G007507234. Robert St. Pierre and Robert Goodrich of Abt Associates offered some valuable feedback on an earlier version of this paper. Paul Williams of Dallas Independent Schools, Bill White of the University of Oregon, and Peter Sharpe of Rustin College (Melbourne) assisted in the data analyses. Robert and Jane Donahue organized computer processing of the data. Harriet Kandelman and Doug Carnine assisted in the editing of the manuscript.

2. Flint, MI, and Flippin, AR, also agreed to participate, but since they were four-year (rather than three-year) programs, it was deemed inappropriate to pool these results with the others (Goodrich & St. Pierre, 1979). Furthermore, sample sizes in Flippin were too small to analyze. (See Becker & Engelmann, 1978, and Becker & Gersten, 1979, for results in these sites, which show a pattern similar to the present data.)

Biographical Resumes

Wesley C. Becker

Associate Dean for Division of Counseling
and Educational Psychology

University of Oregon

Eugene, Oregon 97403

B.A., M.A., Ph.D. Stanford

Education of the Disadvantaged, behavior theory and education, evaluation of instruction.

Area: Division G

Gersten, Russell M.

Research Associate, Follow Through Project

University of Oregon

Division of Special Education

University of Oregon

B.A., Brandeis University

Ph.D., University of Oregon

Special education and compensatory education, program evaluation, assessment.

Area: Division D

Mailing Address:

Follow Through Project
University of Oregon
Trailer 29B
Eugene, OR 97403